

Title: Processing of Czech court decisions

Author: Bohdan Maslowski

Department: Institute of Formal and Applied Linguistics

Supervisor: Mgr. Barbora Vidová Hladká, Ph.D.

Abstract: The objective of this thesis is a comparison of various language processing methods of Czech case-law documents. In particular, the tasks of extraction of information about parties (names, roles, addresses, etc.) and document classification by two criteria, subject and result have been solved. Machine learning methods are evaluated and compared to rule-based approach. For the purpose of training and evaluation of classifiers, a corpus of 400 Czech case-law documents has been created and manually annotated. The thesis includes a web application used for demonstration of the results of different approaches and a tool for running and evaluation of testing scenarios.

Keywords: natural language processing, information extraction, legislative domain, machine learning, rule-based systems